

Aliasing-Free Wideband Beamforming Using Sparse Signal Representation

Zijian Tang, Gerrit Blacquière, and Geert Leus

Abstract—Sparse signal representation (SSR) is considered to be an appealing alternative to classical beamforming for direction-of-arrival (DOA) estimation. For wideband signals, the SSR-based approach constructs steering matrices, referred to as dictionaries in this paper, corresponding to different frequency components of the target signal. However, the SSR-based approach is subject to ambiguity resulting from not only spatial aliasing, just like in classical beamforming, but also from the over-completeness of the dictionary, which is typical to SSR. We show that the ambiguity caused by the over-completeness of the dictionary can be alleviated by using multiple measurement vectors. In addition, by considering the uniform linear array (ULA) structure, we argue that if the target signal contains at least two frequencies, whose absolute difference phrased in wavelengths is larger than twice the array spacing, the spatial aliasing corresponding to these frequencies will be completely distinct. These properties enable us to adapt the existing ℓ_1 algorithms to extract the target DOAs without ambiguity.

Index Terms—Direction-of-arrival estimation, multiple-dictionary, orthogonal matching pursuit, sparse signal representation, spatial aliasing, uniform linear array, wideband beamforming.

I. INTRODUCTION

DOA estimation by means of beamforming can be subject to ambiguities, for instance due to spatial aliasing. For a uniform linear array (ULA), spatial aliasing occurs when the spacing of the ULA is not small enough (the spacing is larger than half the apparent wavelength). On the other hand, a larger spacing, leading to a larger aperture, is often desired to the benefit of angle resolution. Although the aperture can also be enlarged by embedding more channels, this is not always feasible due to hardware and processing costs as well as restrictions in deployment and storage.

In practical sonar signal processing, many underwater objects emit signals with a very wide band or at a very high frequency. For instance, the toothed whales (odontocetes) can emit clicks sweeping from 0 to 20 kHz. Another example, given at the end of this paper, is a diver equipped with an open-circuit breathing set (scuba). The exhaling sound of the diver is so wideband that it can almost be viewed as white [1]. Wideband “aliasing-free” beamforming has been reported in, e.g., [2]–[5]. These works require statistical knowledge to reconstruct

Manuscript received September 01, 2010; revised December 02, 2010 and March 25, 2011; accepted March 25, 2011. Date of publication April 07, 2011; date of current version June 15, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean-Yves Tourneret. This work was supported by the TNO project 032.31642 and in part by NWO-STW under the VICI program (project 10382).

Z. Tang is with the TNO Defence, Security and Safety, 2509 JG, The Hague, the Netherlands, and also with the Delft University of Technology—Fac. EEMCS, 2628 CD Delft, The Netherlands (e-mail: zijian.tang@tno.nl; z.tang@tudelft.nl).

G. Blacquière is with the TNO Defence, Security and Safety, 2509 JG, The Hague, the Netherlands, and also with the Delft University of Technology—Fac. Civil Engineering and Geosciences, 2628 CN Delft, The Netherlands (e-mail: gerrit.blacquiere@tno.nl; g.blacquiere@tudelft.nl).

G. Leus is with the Delft University of Technology—Fac. EEMCS, 2628 CD Delft, The Netherlands (e-mail: g.j.t.leus@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2011.2140108

the signal/noise subspace, which is (for both signal and noise) not always available or reliable due to the volatile underwater acoustic environment. For this reason, the delay-sum beamformer [6] is sometimes preferred in practical sonar applications, although it is more susceptible to spatial aliasing and renders a poor resolution. Most notably, [2]–[5] are not free of spatial aliasing in a strict sense. First, it is often assumed that the array manifolds of the target DOAs are independent for all the frequencies, a situation that lacks theoretical guarantees. Even if this assumption is true, the aliasing-induced ambiguity is not completely resolved: at most, it is suppressed to be indistinguishable from the background noise.

Recently, the sparse signal representation problem, a topic bearing a close affinity with compressive sensing (CS) [7], has attracted enormous attention. The idea of utilizing SSR (and CS) in narrowband DOA estimation has been reported in various contexts. In [8], CS is applied in the time domain to reduce the ADC sampling rate for each channel of the array. A more common approach is to use SSR, as an alternative to conventional beamforming, to improve the angle resolution [9], [10], or to reduce hardware complexity [11]. Like the delay-sum beamformer, the SSR-based methods do not require signal statistics. On the other side, they rely on the assumption that ambiguity is not present. Note that the ambiguity inherent to SSR-based methods comes not only from spatial aliasing, but is also due to the over-completeness of the dictionary. The latter is related to the robustness of an SSR system, which is often characterized by the restricted isometry property (RIP) [7].

This paper gives a rigorous ambiguity analysis for DOA estimation in the context of wideband signals. First, we show that the SSR system will be more robust to nonunique solutions if multiple measurement vectors (MMV) are utilized [12]–[14]. Next, we exploit the property that spatial aliasing is frequency-dependent and show that if there exist at least two frequencies, whose absolute difference phrased in wavelengths is larger than twice the array spacing, the spatial aliasing related to these two frequencies will take a completely different form. This implies that by judiciously choosing the frequencies and constructing the corresponding dictionaries, the resulting signal models will share a common sparse structure, which means that the positions of the nonzero coefficients correspond only to the true target DOAs, although they may have different values across the dictionaries. This fact resembles the joint sparsity model-2 (JSM-2) considered in [15], for which we can utilize a similar solver to jointly reconstruct the sparse signal for multiple dictionaries.

The remainder of the paper is organized as follows. Section II describes the wideband array signal model. We briefly review classical beamforming in Section III and propose an SSR-based DOA estimator in Section IV. Some numerical examples are provided in Section V. Section VI concludes the paper.

Notation: We use upper (lower) bold face letters to denote matrices (column vectors). $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ represent conjugate, transpose and complex conjugate transpose (Hermitian), respectively. Calligraphic letters are used to denote sets. j is reserved for the imaginary unit $\sqrt{-1}$. The rank of a matrix \mathbf{X} is expressed by $\text{rank}(\mathbf{X})$. $\|\cdot\|$ defines the ℓ_2 norm. $x[n]$ stands for the n th entry of the vector \mathbf{x} . Depending on context, we use $|\cdot|$ to represent either the cardinality of a set or the absolute value of a scalar.

II. DATA MODEL

We consider a uniform linear array (ULA) comprised of N channels indexed by $n = 0, \dots, N - 1$, which are equally spaced on a line with spacing d . It receives signals radiated from Q point sources. The sources are supposed to be located in the far field such that we can assume a plane wave arrival. For wideband processing, the signal at

each channel after time-sampling is partitioned into P (possibly overlapping) segments, where for each segment, K frequency subbands are computed by e.g., a filter bank or the discrete Fourier transform (DFT) [16]. Let us use $S_{q,k}(p)$ to denote the k th subband (frequency) coefficient computed for the p th segment of the signal that is radiated from the q th target with $p = 0, \dots, P-1$, $k = 0, \dots, K-1$ and $q = 0, \dots, Q-1$; similarly, we use $y_{n,k}(p)$ to denote the k th subband (frequency) coefficient for the p th segment of the signal received at the n th channel. If the bandwidth of each subband is much smaller than its central frequency f_k , the narrowband assumption applies, which enables the following equality [17]:

$$y_{n,k}(pT) = \sum_{q=0}^{Q-1} e^{j2\pi f_k \frac{d}{c} n \sin \theta_{m_q}} S_{q,k}(pT) \quad (1)$$

where c stands for the propagation velocity of the signal and θ_{m_q} for the DOA of the q th target.¹ In the above equation, we have used the time of arrival at the first channel ($n = 0$) as a reference time and neglected the additive noise term just for the sake of notational ease. The aim of this paper is to estimate the target DOAs $\{\theta_{m_0}, \dots, \theta_{m_{Q-1}}\}$.

If we stack the results from all the channels in a vector, $\mathbf{y}_{k,p} := [y_{0,k}(p), \dots, y_{N-1,k}(p)]^T$, it can be concisely expressed as

$$\mathbf{y}_{k,p} = \sum_{q=0}^{Q-1} \mathbf{a}_{k,m_q} S_{q,k}(p) \quad (2)$$

with

$$\mathbf{a}_{k,m} := [e^{j2\pi f_k \frac{d}{c} 0 \sin \theta_m}, \dots, e^{j2\pi f_k \frac{d}{c} (N-1) \sin \theta_m}]^T \quad (3)$$

also known as the array response vector. Throughout the remainder of the paper, the following assumption will be adopted.

Assumption 1: The array response vectors corresponding to different targets are mutually independent.

III. CLASSICAL BEAMFORMING

As an example of classical beamforming, we consider the delay-sum beamformer in this paper [17]. The angle domain is divided into a grid of M points $\Theta := \{\theta_0, \dots, \theta_{M-1}\}$ and scanned for each possible angle θ_m . This is achieved by weighting the signal received by the n th channel with a corresponding phase shift $e^{-j2\pi f_k \frac{d}{c} n \sin \theta_m}$. The results from all the channels are then summed, which is equivalent to $\mathbf{a}_{k,m}^H \mathbf{y}_{k,p}$. In many sonar applications, a range(time)-bearing(angle) image is desired, which can be made by repeating the above procedure for all the subbands; the outputs are then combined and transformed back into the time domain by means of e.g., an inverse Fourier transform. In the end, the signal for the p th segment at the m th angle in the range-bearing image $I(p, m)$ can be computed as [6]

$$I(p, m) := \left| \frac{1}{N} \sum_{k=0}^{K-1} \mathbf{a}_{k,m}^H \mathbf{y}_{k,p} e^{j2\pi f_k p} \right|^2. \quad (4)$$

$I(p, m)$ can be interpreted as the power of the output of a spatial-temporal filter steered to the direction of θ_m . The DOAs are estimated by seeking those θ_m whose corresponding values in $\sum_{p=0}^{P-1} I(p, m)$ are the largest.

The delay-sum beamformer is subject to spatial aliasing. In the case of a ULA, for a certain frequency f_k and angle θ_m , if the channel

¹Notice that we use m_q here to index the q th target DOA. In the next section, it will become clear that m_q stands for the position of θ_{m_q} within a larger angle set that will be defined later on.

spacing d is larger than half the apparent wavelength $\frac{c}{2f_k \sin \theta_m}$, it is possible to find at least another angle $\theta_{m'}$ such that

$$f_k \frac{d}{c} \sin \theta_m = f_k \frac{d}{c} \sin \theta_{m'} + j \quad (5)$$

where j is an arbitrary integer. As a result, $\mathbf{a}_{k,m} = \mathbf{a}_{k,m'}$, which gives rise to multiple peaks in the range-bearing image $I(p, m)$. The ambiguity due to spatial aliasing aggravates with a higher frequency or larger spacing. The latter is, however, often desired in favor of a higher angle resolution.

IV. DOA ESTIMATION VIA SSR

A. Problem Formulation and Existing Methods

Like in classical beamforming, we divide the whole angle search range into a fine grid $\Theta = \{\theta_0, \dots, \theta_{M-1}\}$, with M being a pre-defined parameter. Each θ_m corresponds to a certain array response vector $\mathbf{a}_{k,m}$ defined in (3), which is dependent on the center frequency f_k . Corresponding to f_k , we construct an $N \times M$ steering matrix $\mathbf{A}_k := [\mathbf{a}_{k,0}, \dots, \mathbf{a}_{k,M-1}]$. In the context of SSR, we will refer to \mathbf{A}_k in the sequel as a dictionary.

For the SSR-based approach, we adopt the following assumption.

Assumption 2: The DOAs of the targets are subsumed by the search grid,² i.e.,

$$\{\theta_{m_0}, \dots, \theta_{m_{Q-1}}\} \in \Theta. \quad (6)$$

With Assumption 2, we are allowed to rewrite (2) in a more general form as

$$\mathbf{y}_{k,p} = \mathbf{A}_k \mathbf{x}_{k,p} \quad (7)$$

where $\mathbf{x}_{k,p}$ stands for an $M \times 1$ vector, which contains all zeros except for the positions whose indexes are contained in the set $\mathcal{T} := \{m_0, \dots, m_{Q-1}\}$. Obviously, $\mathbf{x}_{k,p}[m_q] = S_{q,k}(pT)$. With in total P snapshots available, we can extend (7) to

$$\mathbf{Y}_k = \mathbf{A}_k \mathbf{X}_k \quad (8)$$

where \mathbf{Y}_k is an $N \times P$ matrix $\mathbf{Y}_k := [\mathbf{y}_{k,0}, \dots, \mathbf{y}_{k,P-1}]$ and \mathbf{X}_k is an $M \times P$ matrix similarly defined as \mathbf{U}_k . If we assume that the target DOAs during the span of P snapshots remain unchanged, \mathbf{X}_k will obviously have at most Q rows containing nonzero elements.

Generally speaking, \mathbf{A}_k is an over-complete dictionary, which hampers a straightforward estimation of \mathbf{X}_k . In practice, it is reasonable to assume that the number of targets is very small compared to the size of the dictionary and therefore \mathbf{Y}_k admits a sparse representation. To exploit this property, let us introduce the notation $\mathcal{R}(\mathbf{M})$ to represent the operation that collects the indexes of all the nonzero rows of a matrix \mathbf{M} . Hence, $\mathcal{R}(\mathbf{X}_k) = \mathcal{T}$ and $|\mathcal{R}(\mathbf{X}_k)| = Q$. Armed with such a notation, we can formulate a sparse problem out of (8) as

$$\min_{\mathbf{X}_k} |\mathcal{R}(\hat{\mathbf{X}}_k)|, \text{ subject to } \mathbf{Y}_k = \mathbf{A}_k \hat{\mathbf{X}}_k. \quad (9)$$

Because the columns of \mathbf{X}_k share a common sparsity, (9) falls under the framework of the multiple measurement vectors (MMV) approaches considered in [12]–[14] and the group least absolute shrinkage and selection operator (LASSO) proposed in [19], where various algorithms are given to solve (9).

²If we deviate from this assumption, the effect of any discrepancy can be captured by an additional observation noise term. An alternative is to model the discrepancy explicitly as a disturbance on \mathbf{A}_k [18].

B. Aliasing Suppression

Notice that unlike classical beamforming where the DOAs are sought by steering a beamformer to different potential angles, the SSR-based method given in (9) takes an intermediate step to recover the value of \mathbf{X}_k first and then estimates the DOAs by locating the rows of \mathbf{X}_k that contain dominant entries. To facilitate a robust recovery of \mathbf{X}_k , the restricted isometry property (RIP) [7] should be satisfied for \mathbf{A}_k , or in other words, all subsets of Q columns taken from \mathbf{A}_k need to be nearly orthogonal. However, the RIP does not always hold, which gives rise to nonunique solutions to (9) and thus ambiguity in DOA estimation.

In this paper, we differentiate between two cases of ambiguity. First, if the array spacing d is larger than half the apparent wavelength, it will be possible to find at least two angles θ_m and $\theta_{m'}$, for which (5) holds. Equivalently, the corresponding columns in \mathbf{A}_k will be identical. This effect is widely known as spatial aliasing, which classical beamforming is also subject to. The second cause of ambiguity is inherent to the 'fat' matrix \mathbf{A}_k , where some of its columns, even though not identical, are linearly dependent. We will label this case as algebraic aliasing in this paper to differentiate it from spatial aliasing. Note that algebraic aliasing will not pose any problem in classical beamforming since the latter examines the columns of \mathbf{A}_k individually. Although both spatial aliasing and algebraic aliasing are related to the linear dependency among the columns of \mathbf{A}_k , they demand a separate counter-strategy.

Let us first introduce the following proposition that facilitates the elimination of algebraic aliasing.

Proposition 1: Under Assumption 1, if the number of targets Q and channels N satisfy

$$N > 2Q - \text{rank}(\mathbf{Y}_k) \quad (10)$$

then the SSR-based method in (9) will not suffer from algebraic aliasing.

Proof: Let us first rule out the effect of spatial aliasing by coining a new matrix $\hat{\mathbf{A}}_k$, which is comprised of all the distinct columns of \mathbf{A}_k . Resulting from Assumption 1 which states that the array response vectors corresponding to different targets are linearly independent, we have $\mathbf{Y}_k = \hat{\mathbf{A}}_k \hat{\mathbf{X}}_k$, where $\hat{\mathbf{X}}_k$ is comprised of the corresponding entries of \mathbf{X}_k and $\mathcal{R}(\hat{\mathbf{X}}_k) = Q$. Obviously, algebraic aliasing will not exist if we can find a unique solution $\hat{\mathbf{X}}_k$ satisfying $\mathbf{Y}_k = \hat{\mathbf{A}}_k \hat{\mathbf{X}}_k$. This is only possible according to [14, Theorem 2.4] if the Kruskal-rank³ of $\hat{\mathbf{A}}_k$ is larger than $2Q - \text{rank}(\mathbf{Y}_k)$. At the same time, we understand that $\hat{\mathbf{A}}_k$ is a Vandemonde matrix, whose Kruskal-rank will therefore be equal to its rank which is N [21]. With this, we conclude the proof. ■

Remark 1: Note that although $\hat{\mathbf{X}}_k$ can be unambiguously recovered, the true DOAs are still unknown. This is because the way to generate $\hat{\mathbf{A}}_k$ is not unique and hence there exist multiple mappings between $\hat{\mathbf{X}}_k$ and \mathbf{X}_k .

Remark 2: Proposition 1 underwrites the significance of using multiple measurement vectors. If there is only one snapshot available, i.e., $P = 1$, we have $\text{rank}(\mathbf{Y}_k) = 1$ and hence we can discriminate at most $\frac{N}{2}$ targets. On the other side, because $\text{rank}(\mathbf{Y}_k) \leq \text{rank}(\mathbf{X}_k) \leq Q$, (10) suggests that the number of targets that can be handled is upper-bounded by the number of channels N . It is interesting to note that the same constraint is imposed on classical beamforming but for a different reason.

So far, we have concentrated on the data model for a single frequency f_k . Actually for all the frequencies contained in the target signal, we can follow the same procedures leading to (8), giving rise to different

dictionaries $\mathbf{A}_k \neq \mathbf{A}_l$ if $f_k \neq f_l$. We will next show that using multiple dictionaries enables us to eliminate spatial aliasing. To this end, let us introduce the symbol Γ_k to denote the support of all possible DOA solutions for the k th dictionary:

$$\Gamma_k := \{\mathcal{R}(\hat{\mathbf{X}}_k^{(0)}), \mathcal{R}(\hat{\mathbf{X}}_k^{(1)}), \dots\}, \quad (11)$$

with $\hat{\mathbf{X}}_k^{(n)}$ being one of the solutions of (9).

Spatial aliasing is frequency-dependent, which means that for different center frequencies, the resulting ambiguity will not (completely) overlap. Therefore, we can imagine that if we solve (9) for several center frequencies: f_0, \dots, f_{K-1} and combine the solutions in a judicious way, the ambiguity due to spatial aliasing will at least be reduced. In fact, we can invoke the following theorem to eliminate ambiguity completely.

Theorem 1: With Proposition 1 met, if there exist at least two dictionaries, whose corresponding frequencies, say f_k and f_l , satisfy

$$0 < |f_k - f_l| < \frac{c}{2d}, \quad (12)$$

then the intersection of the solution support related to different dictionaries will contain exclusively the target DOAs, i.e.,

$$\bigcap_k \Gamma_k = \mathcal{T}. \quad (13)$$

Proof: With Proposition 1 satisfied, we can exclude the ambiguity due to algebraic aliasing and need to focus only on spatial aliasing. Let us proceed with a counter-example. Suppose θ_m is one of the target angles and $\theta_{m'} \neq \theta_m$ is spatial aliasing contained in both dictionaries corresponding to f_k and f_l , which implies that $\{\theta_m, \theta_{m'}\}$ belongs to both Γ_k and Γ_l . In accordance with (5), we then have

$$\begin{aligned} f_k \frac{d}{c} \sin \theta_m - f_k \frac{d}{c} \sin \theta_{m'} &= j_1, \\ f_l \frac{d}{c} \sin \theta_m - f_l \frac{d}{c} \sin \theta_{m'} &= j_2, \\ f_k \frac{d}{c} \sin \theta_m - f_k \frac{d}{c} \sin \theta_{m'} - f_l \frac{d}{c} \sin \theta_m + f_l \frac{d}{c} \sin \theta_{m'} &= j_1 - j_2 \\ &= j_3 \end{aligned}$$

where j_1, j_2 , and j_3 are all integers. Notice that because of the port/starboard ambiguity inherent to the ULA, the search space is limited to $|\theta_m| < 90^\circ$ and hence, j_1 and j_2 cannot be zero. Utilizing the trigonometric identities, the above equations can be rewritten as

$$-2f_k \frac{d}{c} \sin \frac{\theta_m - \theta_{m'}}{2} \cos \frac{\theta_m + \theta_{m'}}{2} = j_1 \quad (14)$$

$$-2f_l \frac{d}{c} \sin \frac{\theta_m - \theta_{m'}}{2} \cos \frac{\theta_m + \theta_{m'}}{2} = j_2 \quad (15)$$

$$-2(f_k - f_l) \frac{d}{c} \sin \frac{\theta_m - \theta_{m'}}{2} \cos \frac{\theta_m + \theta_{m'}}{2} = j_3. \quad (16)$$

In light of (12), it is only possible for (16) to hold if the integer $j_3 = j_1 - j_2 = 0$. On the other hand, (14) and (15) suggest that if j_1 and j_2 are not zero, j_1 cannot be equal to j_2 . Therefore, the angle $\theta_{m'}$ cannot be contained simultaneously in Γ_k and Γ_l , which concludes the proof. ■

Remark 3: In classical beamforming, in order to avoid spatial aliasing, it is required that all the frequencies f_k should satisfy $f_k < \frac{c}{2d}$. In this sense, Theorem 1 can be viewed as a relaxed constraint since, instead of avoiding spatial aliasing, it only aims at avoiding the overlap of spatial aliasing corresponding to different frequencies.

Remark 4: It is interesting to note that many classical wideband beamforming works, e.g., [4] and [5], consider the results to be readily "aliasing-free" if (13) is satisfied. However, [4] does not give a rigorous

³The Kruskal-rank of a matrix \mathbf{M} is defined as the largest integer r for which every set of r columns of \mathbf{M} is linearly independent.[20]

formulation of the related choice of frequencies, but simply asserts that the case where (13) is not satisfied “is very unlikely”. In a more specific manner, [5] claims that all wideband signals will not suffer from spatial aliasing, but the wideband signals considered in [5] have a continuous frequency spectrum and thus it is trivial to find two discrete frequencies, which are very close to each other to comply with (12). Compared to [5], Theorem 1 does not only give a more rigorous formulation of the frequency choices, but it is also of practical significance. For instance, we can devise an active system, which emits a sensing signal containing just as few as two frequency components. Further, in many real-life situations, not all the frequencies are usable for DOA estimation. We will give such an example in the next section.

Remark 5: The optimum DOA estimates should be attained by solving (9) using all the frequencies present. In the simulations, however, we will give a real example showing that this is often infeasible. A judicious choice of frequencies can not only prevent spatial aliasing, but also enhance the performance in a noisy environment. An analogous problem is noted in [22], where the choice of frequencies is shown to be significant for Doppler frequency estimation. How to apply a similar approach to DOA estimation remains to be future work.

C. Aliasing-Free SSR Recovery

The analysis in the previous section suggests that we can formulate a multiple-dictionary (MD) joint optimization problem as

$$\begin{aligned} \min_{\hat{\mathbf{X}}_k} & |\mathcal{R}(\hat{\mathbf{X}}_k)|, \text{ for } k = 0, \dots, K-1 \\ \text{subject to} & \mathbf{Y}_k = \mathbf{A}_k \hat{\mathbf{X}}_k, \text{ and} \\ & \mathcal{R}(\hat{\mathbf{X}}_k) = \mathcal{R}(\hat{\mathbf{X}}_l) \text{ for } k \neq l \end{aligned} \quad (17)$$

whose solution will be free of any ambiguity under Theorem 1.

Note that the MD problem $\mathbf{Y}_k = \mathbf{A}_k \mathbf{X}_k$ for $k = 0, \dots, K-1$ has been explored in various works. In its strictest form, the collaborative spectrum sensing problem handled in [23] requires that $\mathbf{X}_k = \mathbf{X}_l$ for $k \neq l$. This model actually reduces to a single-dictionary problem if we stack \mathbf{Y}_k vertically as $\mathbf{Y} := [\mathbf{Y}_0^T, \dots, \mathbf{Y}_{K-1}^T]^T$ such that $\mathbf{Y} = \mathbf{A}\mathbf{X}$ where \mathbf{A} is defined similarly as \mathbf{Y} and $\mathbf{X} = \mathbf{X}_0 \cdots \mathbf{X}_{K-1}$. The multichannel-image problem considered in [24] imposes a less strict constraint on \mathbf{X}_k by requiring $\sum_k f_k(\mathbf{X}_k) = 0$. Here, $f_k(\cdot)$ denotes some sparsity-inducing function which should be differentiable. This property facilitates a dual descent method, which can be viewed as a variant of basis pursuit (BP) [25]. In its loosest form, the problem described in (17) is generalized in [15] as joint sparsity model-2 (JSM-2), where different signals share the same sparsity, but are measured by distinct dictionaries.

In [15], orthogonal matching pursuit (OMP) is utilized to solve the MD problem. Although the algorithm is based on a single measurement vector and defined in the real domain, its extension to the MMV case and complex domain, as considered in this paper, is straightforward.

OMP seeks the solution iteratively. A judicious choice of the stopping criterion can be critical to the performance as shown in [13] and should be tailored to different applications. The following considerations are taken in the choice of the stopping criterion in this paper. First, the problem tackled in this paper is essentially a detection problem: we are in principle not interested in the signal coefficients themselves, but rather in the positions of the nonzero coefficients. Second, in a typical underwater environment, the signal-to-noise ratio (SNR) is usually very low. For instance, for diver detection in a harbour, we would prefer using high-frequency acoustic signals to avoid interference from the busy traffic. However, high frequencies are subject to a much larger attenuation through the water, resulting in an extremely low SNR [26]. Third, the number of targets in a realistic situation is quite small. These considerations suggest that a criterion that halts OMP after a fixed

number, say J , of iterations would suit best. In practice, we predefine J as an upper-bound on the possible number of targets.

A criticism to OMP is that unlike those convex relaxation methods such as BP [27], OMP does not provide the same guarantees on performance, which is phrased as upper-bounds on $\|\mathbf{X}_k - \hat{\mathbf{X}}_k\|^2$. To formulate our MD problem in terms of BP, we have

$$\min_{\hat{\mathbf{X}}_0, \dots, \hat{\mathbf{X}}_{K-1}} f(\hat{\mathbf{X}}_0, \dots, \hat{\mathbf{X}}_{K-1}) + \lambda \sum_{k=0}^{K-1} \|\mathbf{Y}_k - \mathbf{A}_k \hat{\mathbf{X}}_k\|^2 \quad (18)$$

where λ is a parameter controlling the trade-off between the sparsity of \mathbf{X}_k and the error $\|\mathbf{X}_k - \hat{\mathbf{X}}_k\|^2$; $f(\cdot)$ defines some (differentiable) function that enforces the joint sparsity constraint $\mathcal{R}(\hat{\mathbf{X}}_k) = \mathcal{R}(\hat{\mathbf{X}}_l)$ for $k \neq l$. An elegant example of $f(\cdot)$ is given in [28], which exploits the property that when the joint sparsity constraint is satisfied, the number of nonzero rows of the bigger matrix $[\hat{\mathbf{X}}_0, \dots, \hat{\mathbf{X}}_{K-1}]$ will be minimized. Accordingly, the following function is coined:

$$f(\hat{\mathbf{X}}_0, \dots, \hat{\mathbf{X}}_{K-1}) = \sum_{n=0}^{N-1} g_\sigma \left(\|\mathbf{e}_n^T [\hat{\mathbf{X}}_0, \dots, \hat{\mathbf{X}}_{K-1}]\| \right) \quad (19)$$

where \mathbf{e}_n denotes an all-zero vector except for its n th position which is one and

$$g_\sigma(\alpha) = e^{-\frac{\alpha^2}{2\sigma^2}}, \quad (20)$$

with σ denoting some small number. Note that (19) serves as an approximation of the number of nonzero rows of the bigger matrix $[\hat{\mathbf{X}}_0, \dots, \hat{\mathbf{X}}_{K-1}]$ when $\sigma \rightarrow \infty$.

In the numerical examples we show later on, we will stick to OMP for its much faster implementation, which is essential to many practical applications. In practice, the performance degradation of OMP is often limited. Note that an improved version of OMP, referred to as compressive sampling matching pursuit (CoSaMP), has been proposed in [27], which provides a good compromise between complexity and performance with theoretic performance guarantees. Its extension to the MD problem remains to be future work.

V. NUMERICAL EXAMPLES

We will examine the performance of the proposed method using an OMP algorithm, which halts at $J = 5$ iterations. The numerical examples are based on both synthetic data and real data. For both cases, we consider a ULA comprised of $N = 16$ hydrophones, with a spacing of $d = 0.06$ m. The speed of the signal wave is assumed to be $c = 1500 \frac{\text{m}}{\text{s}}$ (underwater acoustic signal).

1) Test Case 1. Synthetic Data: In this test case, we will generate the received signals as a superposition of several harmonics corrupted by additive white Gaussian noise. The signal received by the n th hydrophone at the t th time-instance admits an expression as

$$\sum_{m_q=0}^{Q-1} \sum_{k=0}^{K-1} e^{j\{2\pi f_k(\frac{t}{f_s} - n \frac{d}{c} \sin \theta_{m_q}) + \alpha_{m_q}\}} + \sigma_w w_{n,t} \quad (21)$$

where $f_s = 96$ kHz stands for the sampling frequency; α_{m_q} represents the phase corresponding to the q th target, which is chosen as a uniform random variable in the range $[0, 2\pi]$; $w_{n,t}$ stands for the additive noise having a normal distribution $\mathcal{N}(0, 1)$ and σ_w is a constant satisfying $\text{SNR} = \frac{K}{\sigma_w^2}$. Throughout this test case, we will assume $\text{SNR} = 3$ dB. Finally, f_k stands for the k th central frequency.

We assume that the target signal is present in $P = 100$ snapshots. With the search grid defined as $\Theta = \{-90^\circ, -89.75^\circ, \dots, 90^\circ\}$, each dictionary defined in (8) has a dimension of 16×720 .

In the first numerical example, let us assume that there are two targets whose DOAs are $\{35^\circ, 39^\circ\}$ and utilize two frequencies $f_0 = 25$ kHz

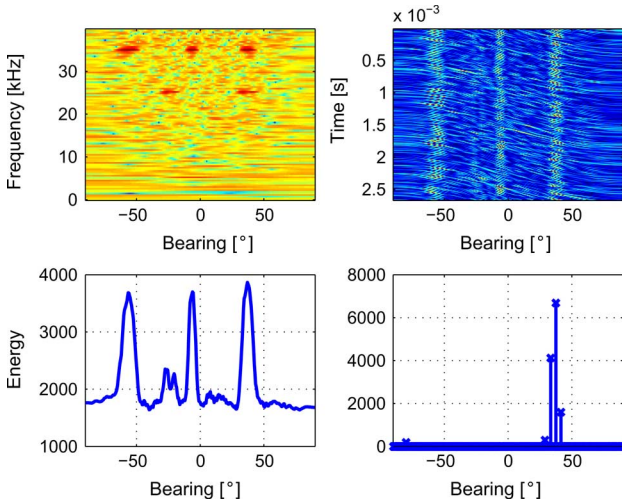


Fig. 1. Comparison of classical beamforming with the proposed method. Upper-left subplot: the frequency-bearing image after beamforming; upper-right subplot: the time-bearing image after beamforming; lower-left subplot: the integrated energy of the time-bearing image; lower-right subplot: the result yielded by the proposed method.

and $f_1 = 35$ kHz to construct the dictionaries. In Fig. 1, we compare the results of the proposed method with classical wideband beamforming. The frequency-bearing image after beamforming is given in the upper-left subplot, where we can see that aliasing arises at different angles for the two frequencies. The time-bearing image is given in the upper-right subplot. The energy-bearing image in the lower-left subplot is attained by integrating the signals along the time axis of this plot. The results yielded by the proposed SSR-based method are given by the lower-right subplot. Compared to the previous subplots, the spatial aliasing disappears. In addition, the two targets seem to be easier to discriminate. Aside from the peaks at the target DOAs, we can also observe three small peaks at other bearings, which correspond to the fact that we have assumed at most five targets to be present and used $J = 5$ as the stopping criterion. Consequently, OMP will in general generate J nonzero coefficients for a noisy system. On the other side, the ambiguity due to overestimating the number of targets is not a serious problem and can be quite straightforwardly removed by setting a proper threshold. Note that thresholding is in practice always indispensable for conventional beamformers.

Fig. 2 indicates the impact of the number of frequencies (dictionaries) utilized on the mean-squared error (MSE) between the estimated and true DOAs. The MSE is calculated as

$$\text{MSE} = \text{E} \left(\sum_{q=0}^{Q-1} (\hat{\theta}_{m_q} - \theta_{m_q})^2 \right) \quad (22)$$

where $\hat{\theta}_{m_q}$ stands for the estimate of θ_{m_q} . The above is evaluated for $Q = 2$ and utilizing 20 possible center frequencies $f_k = 45 - k$ (kHz) for $k = 0, \dots, 19$. Fig. 2 suggests that utilizing more dictionaries increases the estimation precision. The MSE curve flattens off after $K = 10$ due to the fact that during each Monte Carlo run, a pair of random DOAs is generated, which does not necessarily fall on the search grid. When K increases, this kind of modeling error becomes more pronounced, preventing the performance from being further improved. We refer to [18] for a more detailed impact analysis of the modeling error on the SSR problem.

2) *Test Case 2. Real Data:* In this test case, we construct the test data from the raw diver data which were collected during an experiment conducted in June 2009, near Sea Bright, New Jersey, in the U.S. [1].

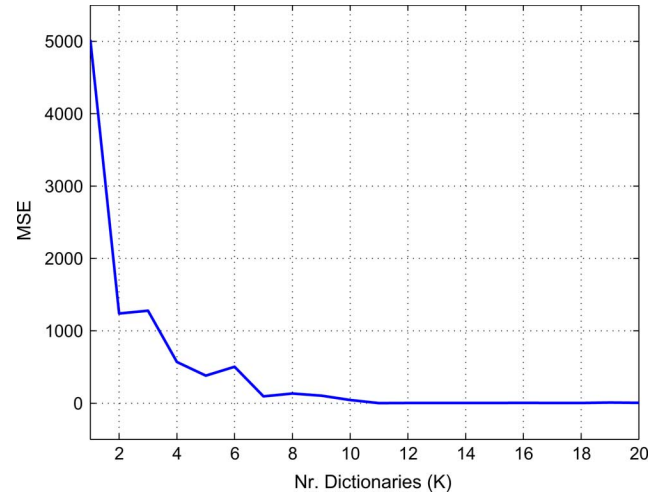


Fig. 2. MSE performance against the number of utilized frequencies.

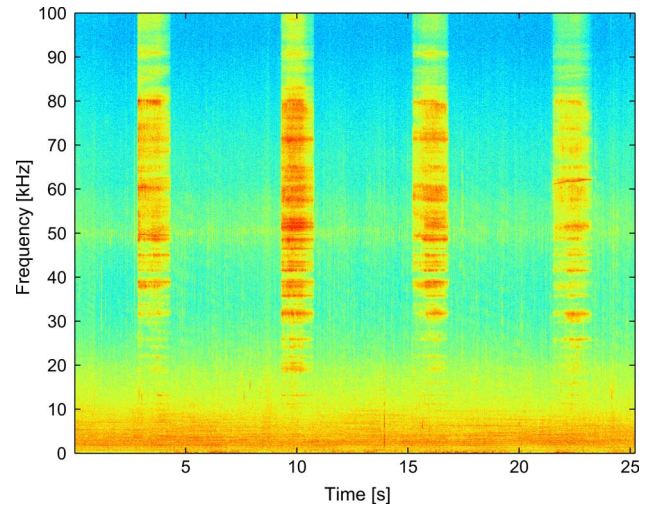


Fig. 3. Spectrogram of the diver.

The diver, at a depth of 10 m, is equipped with an open-circuit breathing set (scuba), and it can be seen in Fig. 3 that the exhaling sound of the diver has a very large bandwidth.

The raw data is actually recorded at a distance of about 10 m from the hydrophone. In the simulation, we plug the raw diver data into a certain underwater propagation model [26], such that we can emulate the effect of different distances and/or directions of the diver. In this example, we generate the received signal as if there are two divers, who are 150 m away from the hydrophone array at angles 52° and 60° , respectively. The results yielded by both classical beamforming and the proposed method are presented in Fig. 4. In the upper-left subplot where the frequency-bearing image is given, we can see that the frequencies lower than 10 kHz are completely useless for DOA estimation: the diver signal is subdued by the ambient noise dominated by the ship traffic in the harbour. In the midfrequency range (between 10 and 12.5 kHz), where the hydrophone array is not subject to aliasing, there is a strong interference signal at a direction around -40° , which possibly comes from a departing ship blowing the horn. The time-bearing image given by the upper-right subplot in Fig. 4 is obtained based on higher frequencies (above 25 kHz), where the spatial aliasing prevents an unambiguous DOA estimate, as in the lower-left subplot. In contrast, the SSR-based method, as shown in the lower-right subplot, renders a correct estimate for the two closely-positioned divers. Note that

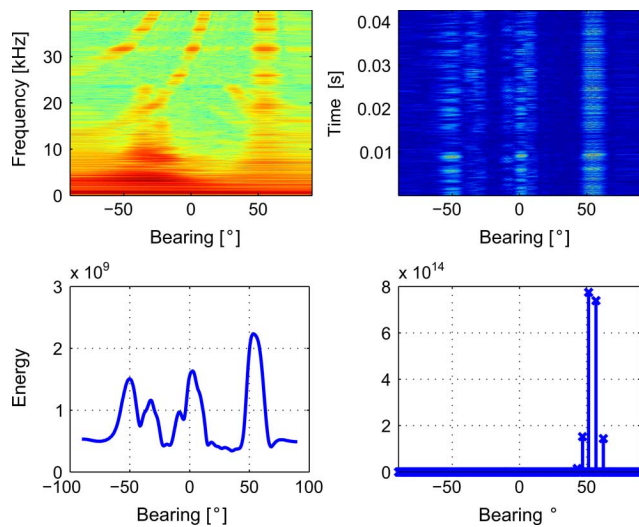


Fig. 4. Comparison of classical beamforming with the proposed method for the diver signal. Upper-left subplot: the frequency-bearing image after beamforming; upper-right subplot: the time-bearing image after beamforming (only signals above 25 kHz are taken); lower-left subplot: the integrated energy of the time-bearing image; lower-right subplot: the result yielded by the proposed method.

for the proposed method we have used the first 100 frequencies above 25 kHz that have the largest amplitude in the frequency response to construct the dictionaries. The frequencies below 25 kHz are not taken into account to minimize the impact of interference. Like in Fig. 1, we observe three minor peaks aside from the two significant peaks at the target DOAs, which is in correspondence with the stopping criterion we have chosen.

VI. CONCLUSION

We applied sparse signal reconstruction for DOA estimation using uniform linear arrays, where an unambiguous decision of the correct angles is hampered by an over-complete dictionary and spatial aliasing. We argued that the ambiguity resulting from an over-complete dictionary can be alleviated by using multiple measurement vectors, while the ambiguity resulting from spatial aliasing can be removed by using multiple dictionaries, each dictionary corresponding to a judiciously chosen frequency.

Based on the above arguments, we formulated the wideband DOA estimation problem as finding the sparse support of the observation signal in the frequency domain. The results shown in the paper were obtained using OMP as the sparse solver, which is attractive due to its fast implementation. For future work, it will be interesting to explore other sparse solvers that are able to render a better trade-off between performance and complexity. In addition to DOA estimation, alternative applications of the multiple-dictionary sparse reconstruction problem will also be investigated.

ACKNOWLEDGMENT

The authors would like to thank Dr. L. Fillinger from the Stevens Institute of Technology for kindly providing the diver data for the simulation.

REFERENCES

[1] B. Borowski, A. Sutin, R. Sutin, H.-S. Roh, and B. Bunin, "Passive acoustic threat detection in estuarine environments," in *Proc. SPIE*, Mar. 2008, vol. 6945.

[2] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wideband sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 823–831, Aug. 1985.

[3] E. D. Di Claudio and R. Parisi, "WAVES: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2179–2191, Oct. 2001.

[4] Y.-S. Yoon, L. M. Kaplan, and J. H. McClellan, "TOPS: New DOA estimator for wideband signals," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 1977–1989, Jun. 2006.

[5] J. Dmochowski, J. Benesty, and S. Affès, "On spatial aliasing in microphone arrays," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1383–1395, Apr. 2009.

[6] R. O. Nielson, *Sonar Signal Processing*. Boston, MA: Artech House, 1991.

[7] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[8] V. Cevher, A. C. Gürbüz, J. H. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation of sparse sources in angle domain," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2008, pp. 2497–2500.

[9] J. J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 702–709, Apr. 2001.

[10] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[11] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," presented at the IEEE Workshop on Statistical Signal Processing, Cardiff, U.K., 2009.

[12] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[13] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[14] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[15] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," 2009 [Online]. Available: <http://arxiv.org/abs/0901.3403>

[16] S. Haykin, *Array Signal Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[17] H. L. van Trees, *Optimum Array Processing (Detection, Estimation and Modulation Theory, Part IV)*. New York: Wiley, 2002.

[18] H. Zhu, G. Leus, and G. B. Giannakis, "Sparse regularized total least-squares for sensing applications," presented at the IEEE Signal Process. Workshop on Signal Process. Adv. Wireless Commun. (SPAWC), Marocco, Jun. 2010.

[19] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. B (Stat. Methodol.)*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[20] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Its Appl.*, vol. 18, no. 2, pp. 95–138, 1977.

[21] N. D. Sidiropoulos and X. Liu, "Identifiability results for blind beamforming in incoherent multipath with small delay spread," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 228–236, Jan. 2001.

[22] A. Ferrari, C. Béranger, and G. Alengrin, "Doppler ambiguity resolution using multiple PRF," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 3, pp. 738–751, Jul. 1997.

[23] Z. Tian, "Compressed wideband sensing in cooperative cognitive radio networks," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2008, pp. 1–5.

[24] N. Ramakrishnan, E. Ertin, and R. L. Moses, "Enhancement of coupled multichannel images using sparsity constraints," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2115–2126, Aug. 2010.

[25] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, Mar. 2006.

[26] R. J. Urick, *Principles of Underwater Sound*. New York: McGraw-Hill, 1975.

[27] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, Apr. 2008.

[28] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell(2,0)$ norm approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4646–4655, Sep. 2010.